



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2011

SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants

Flórez, Lope A ; Gunka, Katrin ; Polania, Rafael ; Tholen, Stefan ; Stülke, Jörg

Abstract: **BACKGROUND:** Several computational methods exist to suggest rational genetic interventions that improve the productivity of industrial strains. Nonetheless, these methods are less effective to predict possible genetic responses of the strain after the intervention. This problem requires a better understanding of potential alternative metabolic and regulatory pathways able to counteract the targeted intervention. **RESULTS:** Here we present SPABBATS, an algorithm based on Boolean satisfiability (SAT) that computes alternative metabolic pathways between input and output species in a reconstructed network. The pathways can be constructed iteratively in order of increasing complexity. SPABBATS allows the accumulation of intermediates in the pathways, which permits discovering pathways missed by most traditional pathway analysis methods. In addition, we provide a proof of concept experiment for the validity of the algorithm. We deleted the genes for the glutamate dehydrogenases of the Gram-positive bacterium *Bacillus subtilis* and isolated suppressor mutant strains able to grow on glutamate as single carbon source. Our SAT approach proposed candidate alternative pathways which were decisive to pinpoint the exact mutation of the suppressor strain. **CONCLUSIONS:** SPABBATS is the first application of SAT techniques to metabolic problems. It is particularly useful for the characterization of metabolic suppressor mutants and can be used in a synthetic biology setting to design new pathways with specific input-output requirements.

DOI: <https://doi.org/10.1186/1752-0509-5-5>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-84622>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 2.0 Generic (CC BY 2.0) License.

Originally published at:

Flórez, Lope A; Gunka, Katrin; Polania, Rafael; Tholen, Stefan; Stülke, Jörg (2011). SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. *BMC Systems Biology*, 5:5.

DOI: <https://doi.org/10.1186/1752-0509-5-5>

RESEARCH ARTICLE

Open Access

SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants

Lope A Flórez¹, Katrin Gunka¹, Rafael Polanía², Stefan Tholen¹, Jörg Stülke^{1*}

Abstract

Background: Several computational methods exist to suggest rational genetic interventions that improve the productivity of industrial strains. Nonetheless, these methods are less effective to predict possible genetic responses of the strain after the intervention. This problem requires a better understanding of potential alternative metabolic and regulatory pathways able to counteract the targeted intervention.

Results: Here we present SPABBATS, an algorithm based on Boolean satisfiability (SAT) that computes alternative metabolic pathways between input and output species in a reconstructed network. The pathways can be constructed iteratively in order of increasing complexity. SPABBATS allows the accumulation of intermediates in the pathways, which permits discovering pathways missed by most traditional pathway analysis methods. In addition, we provide a proof of concept experiment for the validity of the algorithm. We deleted the genes for the glutamate dehydrogenases of the Gram-positive bacterium *Bacillus subtilis* and isolated suppressor mutant strains able to grow on glutamate as single carbon source. Our SAT approach proposed candidate alternative pathways which were decisive to pinpoint the exact mutation of the suppressor strain.

Conclusions: SPABBATS is the first application of SAT techniques to metabolic problems. It is particularly useful for the characterization of metabolic suppressor mutants and can be used in a synthetic biology setting to design new pathways with specific input-output requirements.

Background

A holistic understanding of cellular metabolism is central to systems biology and metabolic engineering: In order to amplify the flux through production pathways in industrial strains we have to understand how the metabolic network responds to our interventions.

Several methods can suggest rational interventions that may lead to favourable industrial phenotypes (see [1] for a review). Their goal is to optimize the distribution of metabolic fluxes towards the product of interest, either directly (e.g. FBA, MOMA or ROOM) or indirectly by coupling it to another characteristic (e.g. Opt-Knock) that facilitates further strain improvements via mutation and screening.

While these methods can predict a final flux distribution, they do not predict the range of genetic and metabolic responses of the organism after the targeted mutation. At the same time, it would be highly desirable to have tools that may predict these responses, since they can suggest ways to generate more stable strains, or accelerate the adaptation to an intended optimal flux. The challenge of the question is the need to understand why particular genetic responses make sense in an evolutionary setting. Thus, the ultimate question is: Which parallel pathways - that were not active previously - result in an adaptive advantage under the screening conditions?

Pathway analysis has received increased attention due to the reconstruction of genome scale metabolic networks for many organisms. These methods can be divided into two categories: stoichiometric and path oriented (see [2] for a review). The first approach generates all pathways that conform to the pseudo-steady-state

* Correspondence: jstuelk@gwdg.de

¹Department of General Microbiology, Georg-August-University of Göttingen, Göttingen, Germany

Full list of author information is available at the end of the article

assumption for internal metabolites. However, it presents two problems: the number of predicted pathways is in the order of millions for genome scale models, making the approach totally intractable for the question at hand [3]. Its second shortcoming is the constraint imposed by the pseudo-steady-state assumption for internal metabolites. This assumption may rule out feasible pathways or (in case we include a large number of “freely available” metabolites) result again in a combinatorial explosion of pathways. The alternative approach - path oriented pathway reconstructions - is advantageous since it usually generates a small (and thus tractable) set of possible pathways. This is due to the choice of starting and ending metabolites and heuristics on the characteristics of the “optimal” pathway. However, the path-oriented approach may result in unrealistic pathways that consume internal metabolites not present in sufficient quantities inside the cell.

What is needed is an algorithm that reconstructs stoichiometrically balanced pathways in increasing order of complexity, with relaxed mass-balance constraints in comparison to the traditional pseudo-steady-state restriction.

A solution based on mixed-integer linear programming (MILP) has been suggested by de Figueiredo *et al.* [4], but it has not been used in an evolutionary context so far. Here we describe the use of Boolean satisfiability (SAT, [5]) for the reconstruction of alternative pathways in metabolic networks. Given a set of basal metabolites (that are considered freely available) and a set of target metabolites (whose concentration *must* increase), our SAT method constructs the shortest pathway between the basal and target sets (SPABBATS) of metabolites that is stoichiometrically balanced, while allowing the concentration of the intermediate metabolites to increase, if needed. The constraints are more relaxed than the ones for e.g. flux balance analysis, thus retaining the metabolically significant pathways. Using the algorithm iteratively, we obtain a prioritized list of pathways, whose elements can be tested individually by common molecular biology techniques.

To demonstrate the power of this concept, we applied the SPABBATS algorithm to a complex physiological problem, which is a result of an evolutionary experiment. We have elucidated a novel pathway of glutamate degradation present in the metabolic network of *B. subtilis* that had been decryptified upon inactivation of the normal glutamate catabolic genes. By using our SAT approach, we proposed four different new pathways that could be present in the mutant to utilize glutamate as single carbon source. These predictions were experimentally tested and revealed that one of these pathways was indeed active in the mutant strain and that this novel “suppressor” pathway is required and sufficient for

glutamate utilization. This proves that the results of our approach correspond to valid metabolic alternatives for living cells.

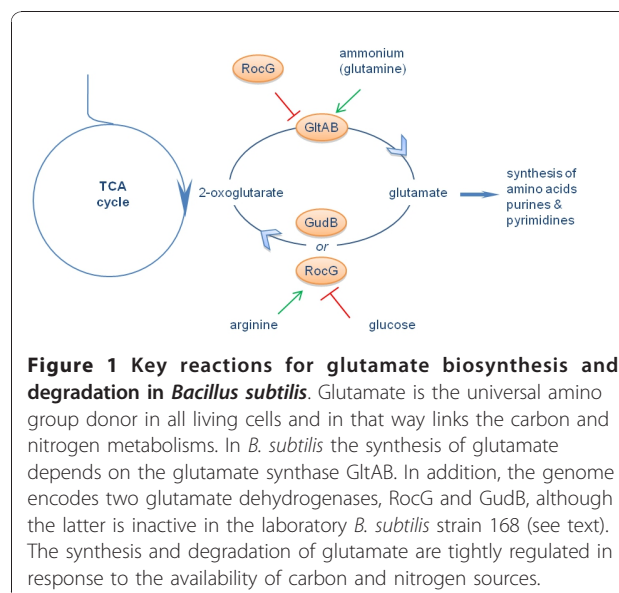
Results

Isolation of a mutation that allows a bypass of the glutamate dehydrogenase for the utilization of glutamate

Glutamate is the most abundant metabolite in a bacterial cell. Although its exact concentration in *B. subtilis* is unknown, it is known to account for about 40% of the internal metabolite pool of an *Escherichia coli* cell [6]. Glutamate serves as an osmotic regulator [7], as well as universal amino group donor in anabolism thus linking carbon and nitrogen metabolism [8]. In *B. subtilis*, at least 37 reactions make use of glutamate as cofactor for transamination [9].

The key reactions of glutamate biosynthesis and degradation in *B. subtilis* are summarized in Figure 1. 2-oxoglutarate, an intermediate of the citric acid cycle, is aminated by the glutamate synthase, encoded by the *gltA* and *gltB* genes. Glutamate degradation to 2-oxoglutarate requires the glutamate dehydrogenase RocG. Additionally, the laboratory strain *B. subtilis* 168 harbours a cryptic gene, *gudB*, coding for an inactive glutamate dehydrogenase. This gene is readily decryptified in *rocG* mutants [10,11]. In addition, RocG controls the expression of the *gltAB* operon and therefore prevents glutamate biosynthesis in the presence of arginine [12,13].

Inactivation of both the *rocG* and the *gudB* gene results in loss of any glutamate dehydrogenase activity and concomitant inability of the bacteria to utilize glutamate [10,11]. The *rocG gudB* double mutant strain GP28 grows poorly on SP medium (an amino acid-rich



medium) due to the accumulation of degradation products of arginine metabolism [14]. However, cultivation of GP28 on SP plates eventually resulted in the isolation of a mutant (GP717) that carries a mutation inactivating the *gluB* gene, encoding a subunit of the glutamate synthase [11]. This mutation leads to glutamate auxotrophy and might therefore prevent the accumulation of intermediates of arginine degradation. We have observed that toxic intermediates of arginine degradation result in poor growth of mutants lacking a functional glutamate dehydrogenase (our unpublished results). If intrinsic glutamate synthesis is blocked by a mutation, such an accumulation of toxic intermediates might be reduced. A careful analysis of the mutant strain revealed that it had acquired the ability to utilize glutamate as the only source of carbon and energy. This might have resulted from a re-activation of the *rocG* or *gudB* genes or from the establishment of a novel pathway for glutamate utilization. We tested therefore the *rocG* and *gudB* alleles by PCR analysis. Both the transposon insertion in *rocG* and the replacement of the *gudB* gene by a chloramphenicol resistance gene were identical to the parent strain GP28. Clearly, a new pathway of glutamate degradation was activated in this suppressor mutant that was not active in the wild type and *rocG gudB* mutant cells.

Development of a pathway-finding algorithm

The most reasonable hypothesis to explain the suppression was that the mutation had activated a redundant pathway that is inactive in the wild type strain in a medium with glutamate as single carbon source. Since glutamate is a highly abundant metabolite and is involved as a substrate in 20 reactions in *B. subtilis*, it was not obvious which mutation could have lead to glutamate utilization proficiency in *B. subtilis* GP717.

To address this problem by use of the power of bioinformatics, we developed an approach that harnesses the strengths of Boolean satisfiability (SAT) to find valid pathways (see Materials and Methods). It is able to find short pathways between a basis and a target set (SPAB-BATS) of metabolites that can operate in a sustained way. It is convenient for its focus on short pathways and the fact that it can calculate pathways that comply with the steady-state constraint. It also allows the relaxation of this constraint, by allowing some metabolites to accumulate if necessary.

The first four pathways suggested by our algorithm are presented in Figure 2. In each case, the first step is a transamination reaction that leads to the production of 2-oxoglutarate. The substrate for transamination is then replenished via the remaining reaction(s) of the pathway. The first pathway (Figure 2A) involves transamination to form alanine and subsequent oxidative deamination of alanine by the alanine dehydrogenase Ald resulting in

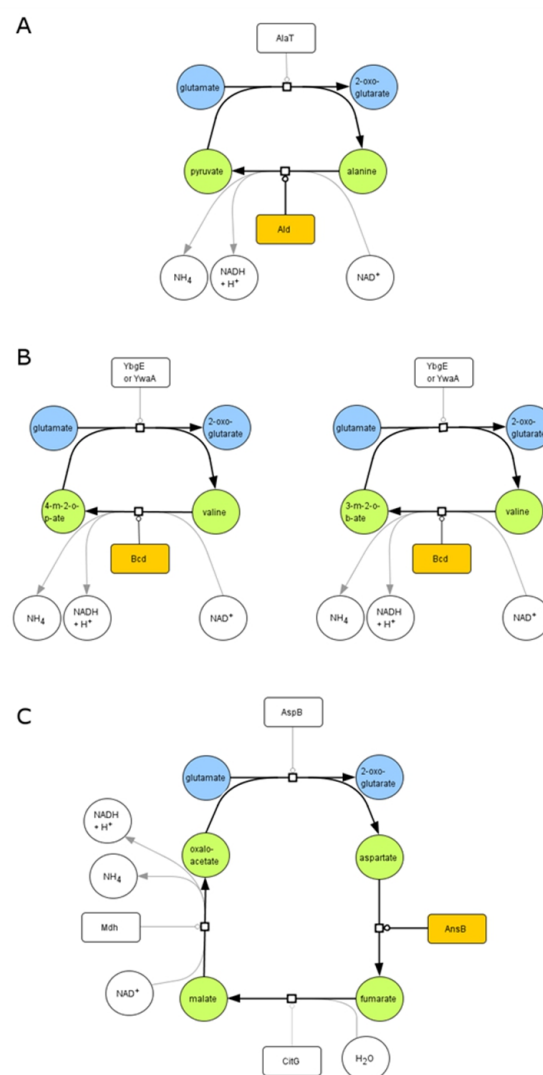


Figure 2 Predictions of alternative pathways for glutamate utilization based on SAT techniques. A *B. subtilis* strain (GP28) was constructed that lacks the glutamate dehydrogenases. An evolutionary adaptation resulted in a strain (GP717) that acquired the capacity to grow on glutamate as single carbon source. Using a SAT based search algorithm (see Materials and Methods) we predicted four alternative pathways that could be activate in the GP717. The genes coding for the enzymes in orange were analyzed further (see text). 4-m-2-o-p-ate = 4-methyl-2-oxo-pentanoate; 3-m-2-o-b-ate = 3-methyl-2-oxo-butanoate.

the net formation of 2-oxoglutarate. The next two pathways (Figure 2B) are very similar and involve enzymes of branched amino acid metabolism. In the transamination step, both pathways use the transaminases YbgE and YwaA. The branched chain amino acid dehydrogenase Bcd is then used for the oxidative deamination of the transamination products valine or leucine. Again, the net result of this pathway is the production of

2-oxoglutarate from glutamate. The last pathway (Figure 2C) requires four steps, (i) the reaction of the aspartate aminotransferase AspB, (ii) the deamination of aspartate to fumarate by the aspartase AnsB, (iii) the fumarase reaction (CitG) of the citric acid cycle, and finally (iv) the oxidation of malate by the malate dehydrogenase Mdh. As described for the other pathways, this reaction sequence results in the net formation of 2-oxoglutarate from glutamate. Since the original mutant GP28 did not grow with glutamate as the single carbon source, it is obviously not able to use any of these proposed pathways suggesting that they were activated by a suppressor mutation in GP717.

Experimental validation of the predictions

Our experiments were performed in minimal medium suggesting that the activity of transaminases was not limiting. Similarly, the two enzymes of the citric acid cycle (CitG and Mdh) are constitutively expressed [15-17]. Thus, the mutation may have affected the expression of one of the deaminases Ald, Bcd or AnsB. This hypothesis was tested by reverse transcription-real-time quantitative PCR. As shown in Figure 3, the levels of *ald* and *bcd* mRNA are comparable for the original mutant GP28 and the suppressor strain GP717. In contrast, a strong increase of the expression of the *ansAB* operon encoding the asparaginase and aspartase was observed for the suppressor mutant that was able to utilize glutamate. This observation suggests that it is the high-level expression of AnsB that allows glutamate utilization in GP717.

The involvement of the aspartase AnsB in the novel glutamate utilization pathway was verified by analysing the effect of a deletion of the *ansAB* operon. Growth of

the original strain GP28, the suppressor mutant GP717 and its isogenic $\Delta ansAB$ mutant derivative GP1154 in minimal medium with glutamate or with glutamate and glucose was recorded. As shown in Figure 4, all three strains were able to grow with glutamate and glucose. In contrast, the deletion of the *ansAB* operon reverted the capability of the suppressor strain of using glutamate as the single carbon source, and the $\Delta ansAB$ mutant GP1154 was unable to grow with glutamate as was the original strain GP28. This finding strongly supports the idea that the activity of the aspartase AnsB is the reason for the ability of the suppressor strain GP717 to utilize glutamate.

The *ansAB* operon is induced in the presence of asparagine due to inactivation of the AnsR repressor [18-20]. A comparative analysis of *ansAB* expression revealed about 30-fold induction by asparagine in GP28, whereas the expression levels were unaffected by the availability of asparagine in the suppressor mutant GP717 (data not shown). The observed induction in the wild type strain is good agreement with previous reports. The loss of regulation in GP717 and the high expression of the operon as compared to GP28 suggest constitutive *ansAB* expression that might be the result of an inactivation of the *ansR* repressor gene.

To test the hypothesis that inactivation of the AnsR repressor allowed glutamate utilization by GP717, we performed two tests: First, we deleted the *ansR* gene of the parental strain GP28 and tested the ability of the resulting strain GP811 to grow with glutamate as the single carbon source. Unlike GP28, this strain GP811 ($\Delta ansR$) grew in CE minimal medium. Thus, inactivation of the *ansR* gene is sufficient to open a new pathway for glutamate catabolism. In a complementary approach, we complemented *B. subtilis* GP717 with a plasmid-borne copy of the *ansR* gene (present on pGP873) and tested the ability of the transformants to use glutamate. While the control strain (GP717 transformed with the empty vector pBQ200) grew well on CE medium, expression of AnsR from the plasmid completely blocked growth in this medium, *i. e.* the utilization of glutamate. This result confirms that a mutation in the *ansR* gene must be present in GP717 and that it is this mutation, which confers the bacteria with the ability to utilize glutamate via the new aspartase pathway.

To identify the mutation in *ansR*, we sequenced the *ansR* alleles of the parental strain GP28 and the glutamate-utilizing suppressor mutant GP717. While the wild type allele of *ansR* was present in GP28, a C-to-A substitution at position 107 of the *ansR* open reading frame was found in GP717. This mutation changes codon 36 from UCA (Ser) to UAA (stop) and results in premature translation termination and the formation of an incomplete and non-functional AnsR repressor protein.

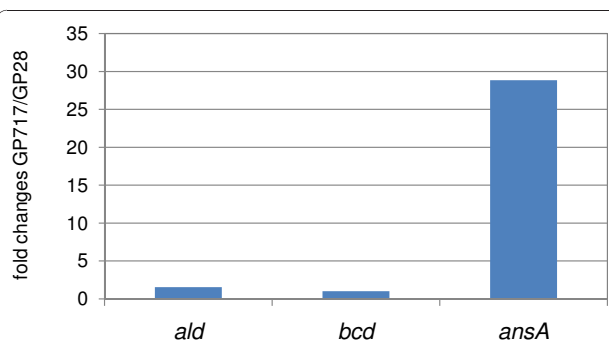


Figure 3 Comparison of gene expression patterns between mutant and parental strains, based on the predictions of the SPABBATS algorithm for pathway analysis. The predictions of the SPABBATS algorithm (see Figure 2) were further characterised by transcription analysis. The expression of the *ald* and *bcd* genes remains constant between the mutant (GP717) and parental (GP28) strains, suggesting that these genes are not involved in the newly activated catabolic pathway. In contrast, the expression of the *ansAB* operon is strongly increased in the mutant. This hints to a gain of function in the mutant strain that was analyzed further.

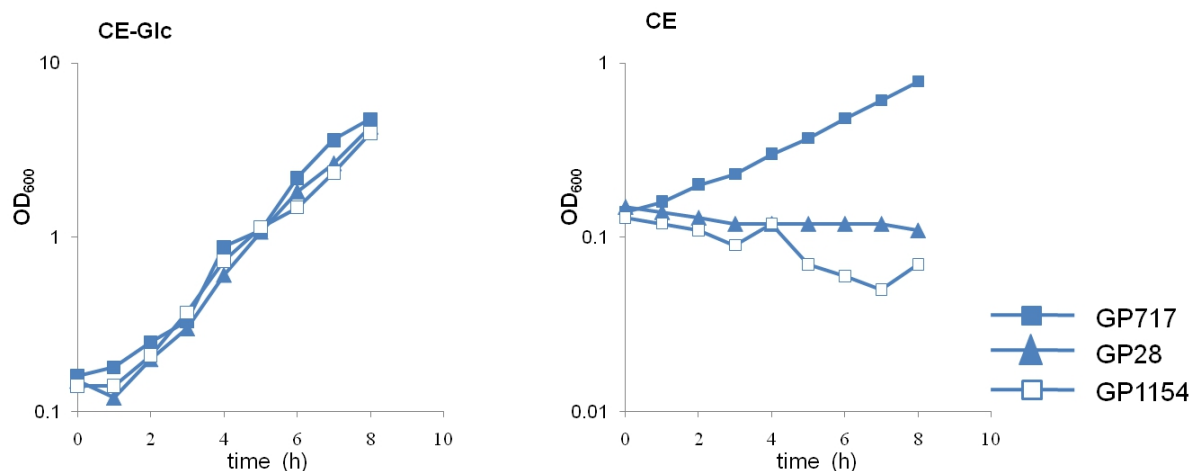


Figure 4 Requirement of the aspartase gene in the alternative pathway for glutamate utilization. The SPABBATS algorithm (see Figure 2) and the transcription analysis (see Figure 3) suggested that the overexpression of the asparaginase and aspartase genes (*ansAB*) is the cause for the metabolic gain of function of the mutant strain GP717. To prove this, the *ansAB* operon was deleted in the GP717 strain. The resulting strain GP1154 lost the capacity to utilize glutamate as single carbon source. This strongly indicates that the induction of the aspartase gene is required and sufficient for the newly activated catabolic pathway. CE = Minimal medium containing 0.5% glutamate, CE-Glc = CE medium with an addition 0.5% glucose.

Taken together, these experiments confirmed that the metabolic pathway predicted by the SPABBATS algorithm corresponds to a valid metabolic state of the *rocG gudB ansR* mutant strain GP717.

Discussion

Comparison of SPABBATS with other methods for metabolic analysis

Flux balance analysis [21] and the majority of methods derived from it are based on constraining the admissible intracellular flux space to steady-state and choosing an adequate optimality criterion to calculate intracellular fluxes. Commonly used optimization criteria are biomass production and the maximization of energy output.

Although these methods predict the essentiality of genes with high accuracy [9], they are less suited for the characterization of alternative metabolic pathways in viable mutants. On the one hand, by restricting the admissible intracellular flux to steady-state, they discard pathways where a by-product accumulates. Nonetheless, the cell is still viable if this by-product is consumed by other pathways in the cell, not directly related to the process that is studied. SPABBATS solves this problem by allowing a larger flux-space, where intermediate products can accumulate, if necessary.

On the other hand, the optimality criterion can be artificial. For instance, maximizing cellular growth might lead to a theoretical maximum growth rate, or a flux

distribution that is as close to the wild-type flux as possible, but it is hard to argue that the regulatory network of the strain is directed to the same target. The pathways discovered by SPABBATS are a structural property of the network and do not depend on an extrinsic optimality criterion (beyond the number of reactions of the resulting pathway). For this reason, the resulting pathways can be interpreted objectively.

Other methods for structural decomposition (e.g. extreme pathways and elementary flux modes, see [2] for a review) rely on the same steady-state restriction of FBA related methods and for this reason share some of their disadvantages. Moreover, SPABBATS does not require the calculation of all possible pathways. Instead, it can be used iteratively to calculate pathways of increasing length, which results in a dramatic improvement in performance for finding relevant pathways in large networks.

An advantage over the method of de Figueiredo *et al.* [4] is that we do not make use of an optimization framework, but select for satisfiability instead. Similar problems in other areas of computational biology (e.g. [22]) show a performance improvement of SAT methods over traditional mixed-integer linear programming methods.

Future perspectives

So far, our analysis of networks using SAT has been restricted to metabolic networks. Nonetheless, since SAT is especially suited for problems that involve Boolean constraints, it is possible to expand the analysis to

regulatory networks. For *B. subtilis*, this implies the reconstruction of the metabolic network together with its regulatory complement. This reconstruction is in progress [23,24].

In parallel, we envision the development of novel SAT solvers that are optimized for the solution of metabolic constraints. This will result in the adoption of SAT based methods for metabolic engineering as well as for the design of synthetic circuits that are able to perform computations in the same way as their silicon-made counterparts [25].

Conclusions

In this contribution we have shown the use of SAT techniques to discover alternative pathways that connect sets of starting and target species. In addition, we provided a proof of concept for the applicability of the algorithm. We started with a complex physiological problem in *B. subtilis*: the need to characterize a suppressor mutation that allowed growth on glutamate without glutamate dehydrogenases. SPABBATS predicted four potential pathways for glutamate utilization that were decisive to suggest target genes for experimentation. These experiments confirmed the validity of the SPABBATS' prediction, closing the cycle between modelling and wet lab experimentation.

SPABBATS relies on Boolean satisfiability (SAT) to construct the metabolic pathways. SAT has been used for the determination of haplotypes from sequenced genotypes [22], the analysis of genome biology networks [26], the understanding of myogenic differentiation [27], and the characterization of steady states of regulatory circuits [28,29]. Here we report the first application of SAT techniques to metabolic problems.

The SPABBATS algorithm was applied here to a specific problem, the analysis of glutamate metabolism in *B. subtilis*. However, the solution strategies are applicable to a broad spectrum of metabolic problems. For instance, SPABBATS can be particularly useful in the characterization of suppressor mutants. Moreover, SPABBATS can also be useful in synthetic biology. Although used here to find pathways in a reconstruction of the metabolism of *B. subtilis*, it is also possible to use a database of enzymes as the starting model. In this way, it can be used to construct synthetic pathways that satisfy specific input-output and mass-balance requirements.

Methods

Algorithm for finding short pathways between a basis and a target set of metabolites (SPABBATS)

Our approach draws inspiration from flux-balance analysis (FBA [21]) in the sense that it searches the flux space of a metabolic network for fluxes that comply

with a set of stoichiometric constraints. The major difference to FBA lies in the optimality criterion; in FBA the value to optimize is the target flux. In our case we change from optimization to satisfiability: we search for a flux that satisfies all the constraints, including a maximum number of allowed reactions.

Another important difference, that is a consequence of satisfiability approach, is that we use two variables for each flux instead of one. The first variable is a positive integer, which is a relative measure of the contribution of that particular flux to the total pathway. The second variable is Boolean and defines whether or not the particular flux takes part in the solution.

As in FBA, we define **S** as the stoichiometric matrix of the network with *n* reactions and *m* compounds. Reversible reactions are split into two unidirectional reactions. We divide the set of compounds into three disjoint sets:

- i) **B** is the set of basis compounds that are considered freely available, either because they are provided in the medium, or because they are "currency metabolites", whose concentration is buffered by the whole system (e.g. ATP, ADP, NADH, etc.)
- ii) **T** is the set of target compounds, the ones constrained to be produced in the pathways of interest
- iii) **I** is the set containing all other compounds, that can be intermediates of the resulting pathway

We use different constraints for each of these sets. The compounds in the set **B** are left unconstrained. For each compound in the set **T**, we write a constraint in the form:

$$\sum_{i=1}^n s_{ij} a_i b_i > 0, \quad (1)$$

where s_{ij} is the stoichiometric coefficient of compound *j* in reaction *i*, and a_i and b_i are the integer and Boolean valued variables of reaction *i*, respectively. These constraints mean that in the solution pathway the overall flux to these metabolites should be positive.

For the compounds in the set **I** we use a constraint similar to (1), with the difference that we use a "greater than or equal to" (\geq) sign. In FBA, an equality sign is used here, to constraint the fluxes to the steady-state space. We purposely do not constrain the pathway to the steady-state space, since the candidate solutions to the problem will not be the only pathway active in the cell and the intermediates that are accumulated in our pathway can be used by other pathways operating in parallel in the system. We require the total flux to these compounds to be non-negative, since the supposition is that they are not present in sufficiently high amounts to allow sustained growth on their consumption.

Next, we add constraints that limit the directionality of reversible reactions. This is done with constraints in the form:

$$b_i + b_j < 2, \quad (2)$$

where b_i and b_j are the Boolean variables of two reactions that together characterize a reversible reaction. These constraints mean that no two directions of a reversible reaction can appear in the final pathway at the same time.

Last, we add a constraint for the total length of the solution. This constraint is:

$$\sum_{i=1}^n b_i \leq k, \quad (3)$$

where k is a positive integer value that determines the maximum number of reactions that can appear in the pathway. This constraint does not immediately find the best solution, but it puts successively stricter upper-bounds to the maximum number of reactions that are allowed. Thus, it is able to find the shortest solution after some iterations by choosing successively smaller numbers for k .

The constraints for the compounds in **T** and **I** are not linear, since each term in the sum is composed of two variables instead of one. For this reason, a linear optimization strategy cannot be used directly. This limitation is not present when we use the SAT-solver HySAT [30]. It is able to find assignments to the variables that satisfy all the constraints in the system, even when these are non-linear. It is also able to detect if no such assignment exists.

If the shortest solution has been found, the best sub-optimal solution can be found by adding an additional constraint in the form:

$$\sum_{i \in K} b_i < k_{op}, \quad (4)$$

where k_{op} is the number of reactions in the shortest solution and K is the set of indices for the reactions in the shortest solution. In other words, we constrain the sum of all the Boolean variables of the optimal solution to be less than k_{op} , thus leaving out the shortest solution

from the solution space. By iterating this process with the Boolean variables of the sub-optimal pathway, we can find solutions with successively higher number of reactions.

The particular implementation of this algorithm for the problem mentioned in the Results section is as follows: we used the genome-scale reconstruction of *B. subtilis* [9]. We removed the biomass “reaction”; it is useful for FBA, since it describes the target flux to cellular growth, but is meaningless in our context. In addition, we removed the reaction “glutamate dehydrogenase” (R_GLUDxi) to simulate the conditions of the strain GP717. We also scaled the non-integer stoichiometric coefficients of the model to integer values (and divided by the greatest common denominator). In our case, the set **B** contained the metabolites ATP, ADP, NAD⁺, NADH, FAD, FADH₂, H₂O, H⁺, NH₄⁺, and glutamate. These “currency metabolites” were chosen due to their participation in most catabolic pathways in the cell. The set **T** contained just 2-oxoglutarate. The remaining compounds were assigned to the set **I**. We set the interval for the a_i to $[[1], 1000]$. The calculations were done using an Intel Core2 Duo processor at 2.66 GHz, with 3.25 GB of RAM. The first pathway (the one involving leucine as intermediate) was found after 28 seconds. All other pathways took less than 8 minutes each to calculate.

Bacterial strains and growth conditions

All *B. subtilis* strains used in this work are derived from the laboratory wild type strain 168. They are listed in Table 1. *E. coli* DH5α [31] was used for cloning experiments. *B. subtilis* was grown in C minimal medium containing ammonium as basic source of nitrogen [32]. Glutamate and/or glucose were added as carbon source as indicated. The medium was supplemented with auxotrophic requirements (at 50 mg/l). *E. coli* was grown in LB medium and transformants were selected on plates containing ampicillin (100 µg/ml). LB, SP and CSE plates were prepared by the addition of 17 g Bacto agar/l (Difco) to LB, SP or CSE medium, respectively.

DNA manipulation and transformation

Transformation of *E. coli* and plasmid DNA extraction were performed using standard procedures [31].

Table 1 *B. subtilis* strains used in this study

Strain	Genotype	Source or Reference
GP28	<i>trpC2 ΔgudB::cat rocG::Tn10 spc amyE::(gltA-lacZ aphA3)</i>	[14]
GP717	<i>trpC2 ΔgudB::cat rocG::Tn10 spc amyE::(gltA-lacZ aphA3) gltB1 ansR-C107A</i>	[11]
GP811	<i>trpC2 ΔgudB::cat rocG::Tn10 spc amyE::(gltA-lacZ aphA3) ΔansR::tet</i>	see Materials and methods
GP1154	<i>trpC2 ΔgudB::cat rocG::Tn10 spc amyE::(gltA-lacZ aphA3) gltB1 ansR-C107A ΔansAB::ermC</i>	see Materials and methods

Restriction enzymes, T4 DNA ligase and DNA polymerases were used as recommended by the manufacturers. DNA fragments were purified from agarose gels using the Nucleospin Extract kit (Macherey and Nagel, Germany). Phusion DNA polymerase was used for the polymerase chain reaction as recommended by the manufacturer. All primer sequences are provided as supplementary material (Additional File 1 Table S1). DNA sequences were determined using the dideoxy chain termination method [31]. All plasmid inserts derived from PCR products were verified by DNA sequencing. Chromosomal DNA of *B. subtilis* was isolated as described [33].

E. coli transformants were selected on LB plates containing ampicillin (100 µg/ml). *B. subtilis* was transformed with plasmid DNA or PCR products according to the two-step protocol described previously [33]. Transformants were selected on SP plates containing tetracyclin (Tet 10 µg/ml), or erythromycin plus lincomycin (Em 2 µg/ml and Lin 25 µg/ml).

Plasmid and mutant strain construction

To express a plasmid-borne *ansR* gene in *B. subtilis*, we constructed plasmid pGP873. For this purpose the *ansR* gene was amplified with the primers KG18 and KG19 using chromosomal DNA of *B. subtilis* as a template. The PCR product was digested with *Bam*HI and *Sal*I and cloned into the overexpression vector pBQ200 [34].

Deletion of the *ansAB* and *ansR* genes was achieved by transformation with PCR products constructed using oligonucleotides to amplify DNA fragments flanking the target genes and an intervening erythromycin and tetracyclin resistance cassettes from plasmids pDG647 and pDG1514, respectively, [35] as described previously [36]. The PCR products were used to transform GP717 and GP28 for the deletion of the *ansAB* and *ansR*, respectively.

Reverse transcription-real-time quantitative PCR

For RNA isolation, the cells were grown to an OD₆₀₀ of 0.5 - 0.8 and harvested. Preparation of total RNA was carried out as described previously [37]. cDNAs were synthesized using the One-Step RT-PCR kit (BioRad) as described [38]. Real time quantitative PCR was carried out on the iCycler instrument (BioRad) following the manufacturer's recommended protocol by using the primers KG26/KG27 for the *ansA* gene, KG38/KG39 for the *ald* gene and KG40/KG41 for the *bcd* gene, respectively. Their recommended data analysis procedure was also used. The *rpsE* and *rpsJ* genes encoding constitutively expressed ribosomal proteins were used as internal controls and were amplified with the primers *rpsE*-RT-*fwd*/*rpsE*-RT-*rev* and *rpsJ*-RT-*fwd*/*rpsJ*-RT-*rev*, respectively. The expression ratios were calculated as fold

changes as described [38]. RT-PCR experiments were performed in duplicate.

Additional material

Additional file 1: Supplementary Table S1. This Table lists the sequences of the oligonucleotide primers used in the experiments.

Acknowledgements

We are grateful to Jens Baumbach for the help with some experiments and to Fabian M. Commichau and Jens J. Landmann for helpful discussions. Christine Diethmaier and Sina Jordan are acknowledged for introducing us to RT-PCR. L. A. F. was supported by the International Molecular Biology Program of the University of Göttingen and the Studienförderwerk Klaus Murmann der Stiftung der Deutschen Wirtschaft. R. P. was supported by a stipend of the Rose Foundation. This work was supported by the Federal Ministry of Education (Research SYSMO network (PtJ-BIO/0313978D)) and the Fonds der Chemischen Industrie to J. S.

Author details

¹Department of General Microbiology, Georg-August-University of Göttingen, Göttingen, Germany. ²Department of Clinical Neurophysiology, Georg-August-University of Göttingen, Göttingen, Germany.

Authors' contributions

LAF, KG, and JS planned the project. LAF and RP conceived and implemented the SPABBATS algorithm. KG and ST provided the experimental validation of the method. LAF, KG, and JS wrote the manuscript and created the Table and Figures. All authors read and approved the final manuscript.

Received: 4 August 2010 Accepted: 11 January 2011

Published: 11 January 2011

References

1. Feist AM, Palsson BØ: The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol* 2008, **26**:659-667.
2. Planes FJ, Beasley JE: A critical examination of stoichiometric and path-finding approaches to metabolic pathways. *Brief Bioinform* 2008, **9**:422-436.
3. Klamt S, Stelling J: Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep* 2002, **29**:233-236.
4. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ: Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 2009, **25**:3158-3165.
5. Claessen K, Een N, Sheeran M, Sörensson N, Voronov A, Åkesson K: SAT-solving in practice, with a tutorial example from supervisory control. *Discrete Event Dyn Syst* 2009, **19**:495-524.
6. Yuan J, Doucette CD, Fowler WU, Feng XJ, Piazza M, Rabitz HA, Wingreen NS, Rabinowitz JD: Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*. *Mol Syst Biol* 2009, **5**:302.
7. Whatmore AM, Chudek JA, Reed RH: The effects of osmotic upshock on the intracellular solute pools of *Bacillus subtilis*. *J Gen Microbiol* 1990, **136**:2527-2535.
8. Commichau FM, Forchhammer K, Stülke J: Regulatory links between carbon and nitrogen metabolism. *Curr Opin Microbiol* 2006, **9**:167-172.
9. Oh YK, Palsson BØ, Park SM, Schilling CH, Mahadevan R: Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* 2007, **282**:28791-28799.
10. Belitsky BR, Sonenshein AL: Role and regulation of *Bacillus subtilis* glutamate dehydrogenase genes. *J Bacteriol* 1998, **180**:6298-6305.
11. Commichau FM, Gunka K, Landmann JJ, Stülke J: Glutamate metabolism in *Bacillus subtilis*: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. *J Bacteriol* 2008, **190**:3557-3564.

12. Commichau FM, Herzberg C, Tripal P, Valerius O, Stülke J: **A regulatory protein-protein interaction governs glutamate biosynthesis in *Bacillus subtilis*: the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC.** *Mol Microbiol* 2007, **65**:642-654.
13. Belitsky BR, Sonenshein AL: **Modulation of activity of *Bacillus subtilis* regulatory proteins GltC and TnrA by glutamate dehydrogenase.** *J Bacteriol* 2004, **186**:3399-3407.
14. Commichau FM, Wacker I, Schleider J, Blencke HM, Reif I, Tripal P, Stülke J: **Characterization of *Bacillus subtilis* mutants with carbon source-independent glutamate biosynthesis.** *J Mol Microbiol Biotechnol* 2007, **12**:106-113.
15. Feavers IM, Price V, Moir A: **The regulation of the fumarase (*citG*) gene of *Bacillus subtilis* 168.** *Mol Gen Genet* 1998, **211**:465-471.
16. Jin S, Sonenshein AL: **Transcriptional regulation of *Bacillus subtilis* citrate synthase genes.** *J Bacteriol* 1994, **176**:4680-4690.
17. Blencke HM, Homuth G, Ludwig H, Mäder U, Hecker M, Stülke J: **Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways.** *Metab Engn* 2003, **5**:133-149.
18. Sun DX, Setlow P: **Cloning, nucleotide sequence, and expression of the *Bacillus subtilis* *ans* operon, which codes for L-asparaginase and L-aspartase.** *J Bacteriol* 1991, **173**:3831-3845.
19. Sun D, Setlow P: **Cloning and nucleotide sequence of the *Bacillus subtilis* *ansR* gene, which encodes a repressor for the *ans* operon coding for L-asparaginase and L-aspartase.** *J Bacteriol* 1993, **175**:2501-2506.
20. Fisher SH, Wray LV Jr: ***Bacillus subtilis* 168 contains two differentially regulated genes encoding L-asparaginase.** *J Bacteriol* 2002, **184**(8):2148-54.
21. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245-248.
22. Graça A, Marques-Silva J, Lynce I, Oliveira AL: **Efficient haplotype inference with pseudo-Boolean optimization.** *Algebraic Biology* Springer Verlag Berlin/Heidelberg; 2007, 125-139.
23. Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, Fromion V: **Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*.** *BMC Syst Biol* 2008, **2**:20.
24. Lammers CR, Flórez LA, Schmeisky AG, Roppel SF, Mäder U, Hamoen L, Stülke J: **Connecting parts with processes: SubtiWiki and SubtiPathways integrate gene and pathway annotation for *Bacillus subtilis*.** *Microbiology* 2010, **156**:849-859.
25. Lou C, Liu X, Ni M, Huang Y, Huang Q, Huang L, Jiang L, Lu D, Wang M, Liu C, Chen D, Chen C, Chen X, Yang L, Ma H, Chen J, Ouyang Q: **Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch.** *Mol Syst Biol* 2010, **6**:350.
26. Chin G Jr, Chavarria DG, Nakamura GC, Sofia HJ: **BioGraphE: high-performance bionetwork analysis using the Biological Graph Environment.** *BMC Bioinformatics* 2008, **9**:S6.
27. Piran R, Halperin E, Guttmann-Raviv N, Keinan E, Reshef R: **Algorithm of myogenic differentiation in higher-order organisms.** *Development* 2009, **136**:3831-3840.
28. Tiwari A, Talcott C, Knapp M, Lincoln P, Laderoute K: **Analyzing Pathways Using SAT-Based Approaches.** *Algebraic Biology* Springer Verlag Berlin/Heidelberg; 2007, 155-169.
29. de Jong H, Page M: **Search for steady states of piecewise-linear differential equation models of genetic regulatory networks.** *IEEE/ACM Trans Comput Biol Bioinform* 2008, **5**:208-22.
30. Fränzle M, Herde C, Teige T: **Efficient solving of large non-linear arithmetic constraint systems with complex Boolean structure.** *Journal on Satisfiability* 2007, **1**:209-236.
31. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: a laboratory manual*. 2 edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; 1989.
32. Wacker I, Ludwig H, Reif I, Blencke HM, Detsch C, Stülke J: **The regulatory link between carbon and nitrogen metabolism in *Bacillus subtilis*: regulation of the *gltAB* operon by the catabolite control protein CcpA.** *Microbiology* 2003, **149**:3001-3009.
33. Kunst F, Rapoport G: **Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*.** *J Bacteriol* 1995, **177**:2403-2407.
34. Martin-Verstraete I, Débarbouillé M, Klier A, Rapoport G: **Interaction of wild-type truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon.** *J Mol Biol* 1994, **241**:178-192.
35. Guérout-Fleury AM, Shazand K, Frandsen N, Stragier P: **Antibiotic-resistance cassettes for *Bacillus subtilis*.** *Gene* 1995, **167**:335-336.
36. Wach A: **PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *Saccharomyces cerevisiae*.** *Yeast* 1996, **12**:259-265.
37. Ludwig H, Homuth G, Schmalisch M, Dyka FM, Hecker M, Stülke J: **Transcription of glycolytic genes and operons in *Bacillus subtilis*: Evidence for the presence of multiple levels of control of the *gapA* operon.** *Mol Microbiol* 2001, **41**:409-422.
38. Rietkötter E, Hoyer D, Mascher T: **Bacitracin sensing in *Bacillus subtilis*.** *Mol Microbiol* 2008, **68**:768-785.

doi:10.1186/1752-0509-5-5

Cite this article as: Flórez et al.: SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. *BMC Systems Biology* 2011 **5**:5.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

